DOCUMENT RESUME

ED 340 780 TM 018 036

AUTHOR Cuban, Larry
TITLE The Misuse of Tests in Education.
SPONS AGENCY Congress of the U.S., Washington, D.C. Office of Technology Assessment.
PUB DATE Sep 91
NOTE 14p.; Contractor report prepared for the Office of Technology Assessment titled "Testing in American Schools: Asking the Right Questions." For related document, see TM 018 025.
PUB TYPE Viewpoints (Opinion/Position Papers, Essays, etc.) (120)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Academic Achievement; *Educational Assessment; Educational History; Educational Policy; Elementary Secondary Education; National Programs; *Policy Formation; Political Influences; Student Evaluation; *Testing Problems; Testing Programs; Test Results; Test Use

ABSTRACT

Test misuse is neither isolated nor recent. It is a problem that cannot be easily solved. While test misuse may be reduced or managed, it cannot be eliminated. Test misuse has cut across America's social, economic, and political institutions, including schools. The most flagrant abuse of a test is what happens to the results of the Scholastic Aptitude Test, which has not been designed to evaluate schools or teachers, but which is commonly used for these purposes. Much test misuse stems from media-induced hypersensitivity to student performance. Historic and social factors explain why policymakers and administrators under pressure from public officials and angry citizens slipped into using tests improperly. Two negative consequences have been the use of tests by policymakers as remote-control devices to alter instruction and the spread of test-score pollution, the growing meaninglessness of test scores. Several specific suggestions to alleviate, but not eliminate the problem of test misuse, are: (1) recognize that test abuse is a response to dilemmas in the public schools; (2) abolish policies mandating particular tests; (3) reject proposals for national examinations such as those called for in America 2000; and (4) provide funds to develop and pilot unorthodox tests designed to have students demonstrate understanding through actual performance. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED340780

THE MISUSE OF TESTS IN EDUCATION

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

by
Larry Cuban
Stanford University

for
Office of Technology Assessment
U.S. Congress

September 1991

TM018036

BEST COPY AVAILABLE

THE MISUSE OF TESTS IN EDUCATION

Test misuse is neither isolated nor recent; it is pervasive and historic. Nor is test abuse a technical problem that can be easily solved; it is anchored in intractable, messy dilemmas that have faced public schools in the United States for over a century. Test misuse might possibly be reduced, even managed, but not eliminated.

By test misuse I mean simply that users of multiple-choice, standardized achievement and intelligence tests wittingly or unwittingly ignore the explicit purposes of the test and cautions offered by the test-makers (errors in measurement, for example) and use the results to serve other purposes. Such misuse by policymakers, administrators, and practitioners is pervasive. It is common not only to schools but also to many other social institutions.

In medical care, for example, many doctors routinely order tests to avoid potential malpractice suits. The tests are either redundant--the doctor already knows what the diagnosis is--or marginally unrelated to the patient's condition. Estimates of such test misuse run to almost $15,000,000,000 a year. In both the private and public sectors, employers have used tests that bear little relationship to a job's requirements. Such pre-employment tests have screened out capable minority and women applicants. Courts have ordered police and fire departments in cities, for example, to use other tests of fitness for employment that are more closely linked to the work performed. Or consider the results of blood tests to determine if an employee is HIV-positive. Results have been used to deny white- and blue-collar workers their

insurance coverage, to discriminate in work assignments, and to fire those who have contracted the virus. Finally, for decades until they were ruled illegal, most southern states used literacy tests to deprive African-American voters of their right to vote. Evidence of test abuse cuts across American's social, economic, and political institutions, including schools.

Test misuse in schools. Test-makers have warned repeatedly that using intelligence and achievement tests to screen children for admission to a nursery school or retain five year-olds for another year is violating the purposes of these tests (to provide information to teachers to help plan instruction for students, for example). Yet intelligence tests are given to three and four-year olds to rank candidates for entry into private nursery schools; children in the last few months of kindergarten take tests which will determine who will be retained, who will move into first grade, and who will go to a junior-primary class or some other special class for those not yet ready for first grade.

The most flagrant abuse of a test is what happens to the results of the Scholastic Aptitude Test (SAT). The Educational Testing Service (ETS) has continually alerted users, the media, and policymakers that the test has been designed to predict a student's academic success in college. Because only a portion of each school's student body take the test and because the test does not measure what has been taught in a school, ETS explicitly states that the SAT is not to be used to either determine whether schools are successful in educating their students or rank schools on their academic performance. Nonetheless, hundreds of school boards and

superintendents, scores of legislators, and federal officials publicly proclaim credit for one-point increases in scores and blame television and parents for three-point decreases. Three U.S. Secretaries of Education, fully aware of the purposes of this test and these warnings, have used the SATs and similar tests in ranking the 50 states' scores on what has come to be known as the "wall chart."

Even in the face of test-makers' warnings that standardized achievement test scores for individual students should not be used to either monitor academic performance or rank teachers and schools, various cities and states have aggregated test scores by classroom, school, and district to allocate salary increases, administer penalties to teachers and principals, and determine if schools are academically bankrupt to require removal of their staffs.

Some test abuses have become so blatant and harmful to individual students that courts have ruled against using particular test scores. In the 1970s, for example, giving I.Q. tests in California was prohibited because they discriminated against minority children. Similarly, in Florida after the introduction of minimum competency tests (MCT), many African-American children were denied their diplomas even after completing the necessary requirements for graduation because they had failed the MCT. In the late 1970s, the courts ruled that these tests did not reflect what the students had been taught in high school and therefore could not be used to withhold a diploma from those who had failed the test.

History of misuse. Instances of test misuse in schools are not isolated to the present; they have a long history. With the introduction of mass testing in the schools just after World War I, test-makers had converted scores of Army draftees to a mental-age scale and reported the average mental age for white draftees to be 13. Because psychologists had defined a moron as anyone with a mental age of 7 to 12 years, journalists couldn't resist the punch-line: almost half of the white soldiers who were drafted were classified as morons. Academics declared that spending money on education and improved health was foolish because it allowed weaker individuals to survive. The racism directed at southern and eastern European immigrants found a home in schools using the brand-new intelligence tests. Administrators eager to provide classes that would permit the most able students to move swiftly through the curriculum and the dullest to move at their pace unembarrassed by the remarks of sharper classmates-- tested every student. Believing that these new intelligence tests were accurate indicators of innate intelligence, policymakers' and administrators' racist beliefs about the intelligence of different immigrant groups found a safe home in the test scores of immigrant children. Italians, Polish, Russian, and Hungarian children scored low on these tests and were shunted into special classes while native-born American students were placed elsewhere.

Why has such test abuse been and continues to be so pervasive within schools and across American institutions? One answer is that tests designed by experts carry within them values highly-esteemed in American culture: scientific objectivity, fairness,

competitiveness, and efficiency. Standardized achievement and intelligence tests are products of science and that knowledge is linked to improved health and a high standard of living; tests are fair thus allowing individual merit, not family background to emerge; in tests, anyone who has a pencil gets an equal chance to compete; finally, test results can be gotten cheaply and can be easily reduced to a simple number. With these highly-prized public values, tests get placed on a pedestal. Although this answer may help to explain the exaggerated importance that tests assume in this culture, it does not explain frequent or persistent misuse. What is missing in the answer is the entangled interaction between testing companies, the media, and public pressure for schools to be publicly accountable for student performance.

The abiding faith in public schools as a super-glue binding together disparate groups into a cohesive nation began to decay after World War II. Erosion of that faith accelerated sharply in the late 1950s when education, another Cold War weapon drafted to combat Soviet supremacy, came under severe attack, deepened considerably in the 1960s as the civil rights movement revealed dismaying inequities in the schooling that African-American children received, and deteriorated further in the 1970s and 1980s when commission reports, magazine specials, and television documentaries displayed the supposed failures of public schools. By the early 1990s, constant criticism of the school's failure to remedy knotty social and political problems had gouged deep holes in the faith that public schools were essential to binding a nation together.

A shrinking faith in schools to heal national fractures and solve social problems, of course, was part of the larger skepticism about American institutions that grew throughout the 1960s and 1970s from assassinations of public heroes, a devastating loss of a war that unnerved the nation, and an American President who proclaimed that he was not a crook. By the mid-1970s, the skepticism had hardened into an anti-government bias.

Within this sour climate of skepticism, public schooling as a service rendered by local government would naturally come under increased and intense scrutiny. How do we know that our tax dollars are being spent well? Where *does* all that money go? Calls for schools to account publicly for what they do with students coincided with the expanded use of standardized achievement test scores as a measure of school productivity. By the late 1970s, the publishing of school-by-school test scores in newspapers and by districts themselves had become standard practice in big cities as a way of demonstrating school performance to a public hungry for evidence of high performers displayed in simple, clear information like, for example, in pitching and batting statistics on the sports page.

The media, particularly newspapers at first, played a crucial role in translating the skepticism into concrete stories about school performance. Newspapers, magazines, and television editors and journalists sense what the public will respond to as news and then convert raw data and inaccessible research findings into understandable prose, pictures, and statistics. The imperatives within media to highlight the controversial and sharpen any conflict within a situation easily led journalists into publishing portions of

scholarship that hit readers between the eyes: the Coleman Report (1966) and *Inequality* (1972) led to crisp headlines and television reports that schooling makes little difference in either the academic careers or future work experiences of students; the reportorial hullabaloo over Arthur Jensen's research (1969) and Robert Herrnstein (1971) underscored the centrality of intelligence testing; and a decade later when *A Nation at Risk* (1983) was issued, the media went into a feeding frenzy over the dismal failure of public schooling.

Within this media-induced hyper-sensitivity to student performance the role played by commercial publishers of tests surfaces. Testing, after all, is a profitable business. Revenues from the sale of screening, readiness, and achievement tests, scoring services, and data reports have soared in the last quarter-century. The National Commission on Testing and Public Policy estimated in 1990 that taxpayers spend $100 million per year in buying and scoring state and local tests from test publishers. If the related services that publishers offer (preparation materials, test-item analysis, printed out individual reports, etc.) are added in, the Commission raised the bill taxpayers pay to a half-billion dollars. With such high revenues, it comes as no surprise that test-makers' calls for proper use of their tests get drowned out by the noise of cash registers or get reduced to tiny print in contracts. Even worse is that a few firms make misleading, even false, claims for what a test could do for a beleaguered school system (identify at-risk three year-olds, potential dropouts, etc.); the claims end up in mailboxes of superintendents and legislators.

Here, then, is why tests get abused by policymakers, practitioners, and citizens eager to improve schools. Begin with the highly-prized cultural values embedded in expert-designed tests. Tests are good because they are believed to be scientific, fair to anyone who takes them, encourage healthy competition, and, moreover, are inexpensive tools that can produce believable numbers to aid decisionmaking. Then take the last quarter-century's events as interpreted and mediated by journalists which helped produce an anti-government mood. This sour mood heightened the value of tests as a simple and powerful tool for making schools, colleges and other public institutions accountable to taxpayers. Finally, test publishers saw their market expand enormously in a few decades and acted as other American entrepreneurs would in a similar situation. Taken altogether, these factors explain why well-intentioned policymakers and administrators committed to school improvement but intensely pressed by public officials and angry citizens to demonstrate improvement slipped into using test improperly.

Negative consequences. What the above explanation omits, however, is the chain of negative consequences that spill into classrooms from policymakers and administrators improperly using tests. At least two outcomes of test abuse have become obvious in the last decade: Policymakers use tests as remote control devices to alter instruction; and the spread of test-score pollution.

Many federal, state, and district policymakers have adopted particular tests to drive the curriculum and change how teachers teach. The premise is that if certain items can be inserted into tests and if these tests have high stakes attached to them (allocation of

funds, recognition of high performance, removal of staff for low performance), teachers will alter what they teach and how they teach in order to get high scores on the tests. Moreover, evidence piles up that teachers concentrate on what content and skills will be on the tests. Untested content (e.g., arts, science, etc.) gets neglected. Seen as a cheap way of reforming school and classroom practices, this remote control of local practice from policymakers' desks represents the latest evidence of test abuse and its largely negative consequences..

"Test score pollution," a phrase invented by scholars to describe the growing meaninglessness of test scores, is another consequence of test abuse. Suppose, for example, that standardized test scores rise because teachers have students practice with questions similar to ones that will be on the test, or administrators and teachers clean up answer sheets by erasing stray marks or darkening lightly penciled-in answers, or create a curriculum that matches the skills on the test, or school boards buy commercial materials aimed at improving students' performance, or, as in some instances, teachers actually give students the items that will be on the test. Scholars have found that such practices vary from district to district. Whether regarded as ethical or unethical, these practices have, indeed, increased test scores. Such inflated scores then cannot be interpreted as meaning that students have learned more or can perform the academic skills. It only means that the scores are higher because efforts have been made to raise the scores. Raising the score is the goal, not students learning more. As printing more and more currency that has no gold behind it makes paper money

worthless, polluted scores mock the values supposedly embedded in these scientific, fair, competitive, and efficient tools.

What can be done? Abolish tests? No. Certain tests designed for specific uses, carefully administered and with results interpreted cautiously can serve well the different interests of students, practitioners, and policymakers. But with the intersecting factors that I identified earlier (public insistence that schools be accountable for high academic performance, the role of test publishers, and the media), these caveats often get ignored. Nonetheless, standardized achievement tests are here to stay. So what should be done?

What emerges from this examination of test abuse are a few intractable but very familiar dilemmas facing American public schools: How can policymakers and practitioners provide an equal and efficient schooling that cultivates each individual's potential for masses of children who have diverse abilities, varied attitudes toward learning, and unequal motivation? How can policymakers who need sustained public support for schools and utterly depend upon practitioners for doing the daily work with students maintain credibility with both constituencies and still display evidence of satisfactory performance in schools? The conflicting values within each of these dilemmas suggest that compromises must be made since limited time and money prevent fully satisfying any particular value.

Multiple-choice, standardized intelligence and achievement tests and their documented misuse have been an instance of trying to trade-off conflicting values, of trying to reconcile competing

choices. As the abuses pile up and unintended consequences become painfully evident, new ways of balancing conflicting values need to be found. The situation is not a technical problem that can be solved by more information to parents, a better multiple-choice test, or better trained staff. The situation is a high-stakes dilemma that is invulnerable to a technical solution. Dilemmas, however, can be managed, certainly better than they have been. But how?

Numerous technical suggestions have been made to reduce test abuse and its consequences. For example, some critics urge more careful administration of tests by state officials and more security for the actual tests prior to their being given to teachers and students. Others have suggested a political solution such as a national agency that monitors test design, administration, and interpretation of results--a Consumers Union for testing. These suggestions are sensible and will help. They do, however, nibble at the edges of dilemma and do not reconcile the core conflict between competing values. Hence, I offer a few suggestions for federal and state policymakers that confront the dilemmas I identified.

Suggestions. *Recognize that test abuse is basically a response to inherent and historic dilemmas in public schools.* Such recognition is a start that might prod federal and state policymakers to move away from the simplistic notion of finding just the right test to combine measuring individual student's grasp of content and skills, monitoring school and district performance, and holding districts accountable for how they perform. Such a quick, cheap technical solution does not exist on this planet. Nor does such a test solve these dilemmas.

Abolish policies mandating particular tests. Much test abuse occurs because legislators and other policymakers often seek to reform schools and make them accountable by requiring students to take particular tests designed for distinctly other purposes. Finding different tests that match legislative purposes with educational ones might offer promising outcomes or, better yet, finding other ways than using tests to improve schools and make them accountable for what they do.

Reject the recent proposals of President George Bush in America 2000 for national exams (called "American Achievement Tests") composed of multiple-choice questions given to students to determine not only individual, school, and district progress in academics but also to allocate federal funds. Without altering any of the conditions that I identified earlier such a national test would only perpetuate further misuse of a test and worse consequences for students and teachers than already exist.

Provide funds to develop and pilot unorthodox tests designed to help students demonstrate understanding through actual performance. Such tests, some of which do exist in various cities and states, can be then made available to other districts across the country. Such alternatives would help reduce the misuse of tests.

These modest suggestions will disappoint policymakers seeking the grand, simple recommendation that sweeps away the pervasive and historic practices of test abuse and its negative consequences. Sadly, there are no such solutions. There are only better ways of managing dilemmas that just won't go away.